

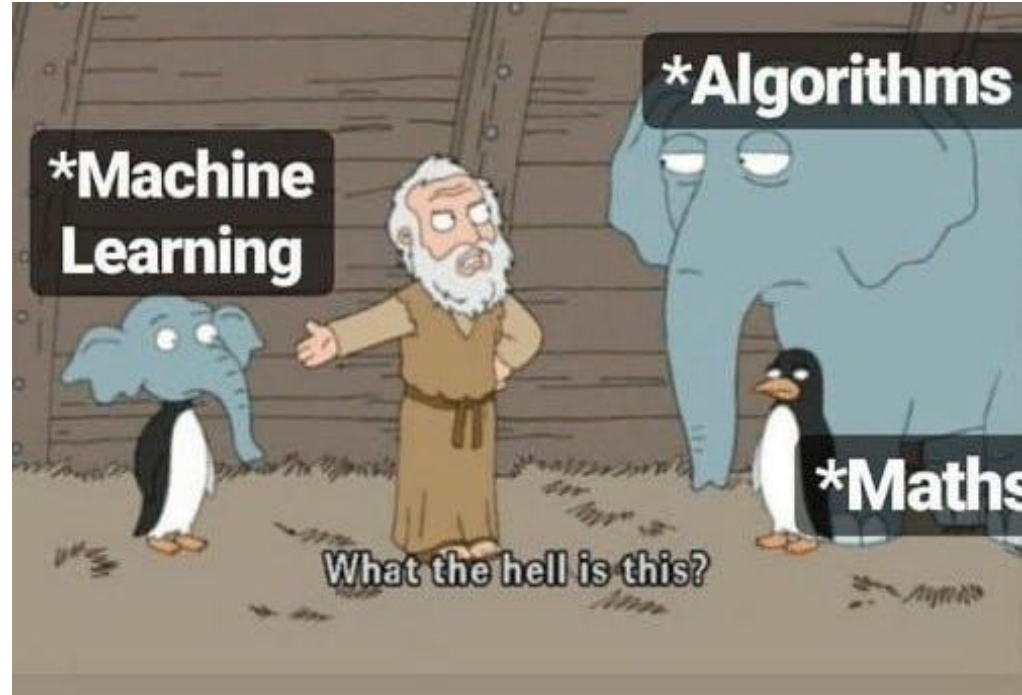
Cierre Aprendizaje Automático

2do cuatri 2023
Paula Pérez Bianchi



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA



Qué harían si necesitamos...

<https://www.menti.com/alqv2bbn1ymp>

ALGORITMOS CLÁSICOS

- Ordenar un arreglo
- Buscar un elemento en una lista
- Multiplicar matrices
- Dados 100.000.000 de tweets, contar los que nombran a Taylor Swift



Y si ahora queremos...

→ Dado un dataset con datos de perros y gatos, quiero identificar los gatos

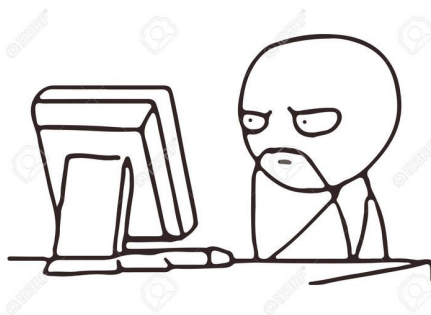
ALGORITMOS CLÁSICOS



vs



¿Qué hacemos?



Hay problemas con reglas tan complejas que no podemos atacarlos con programación clásica...

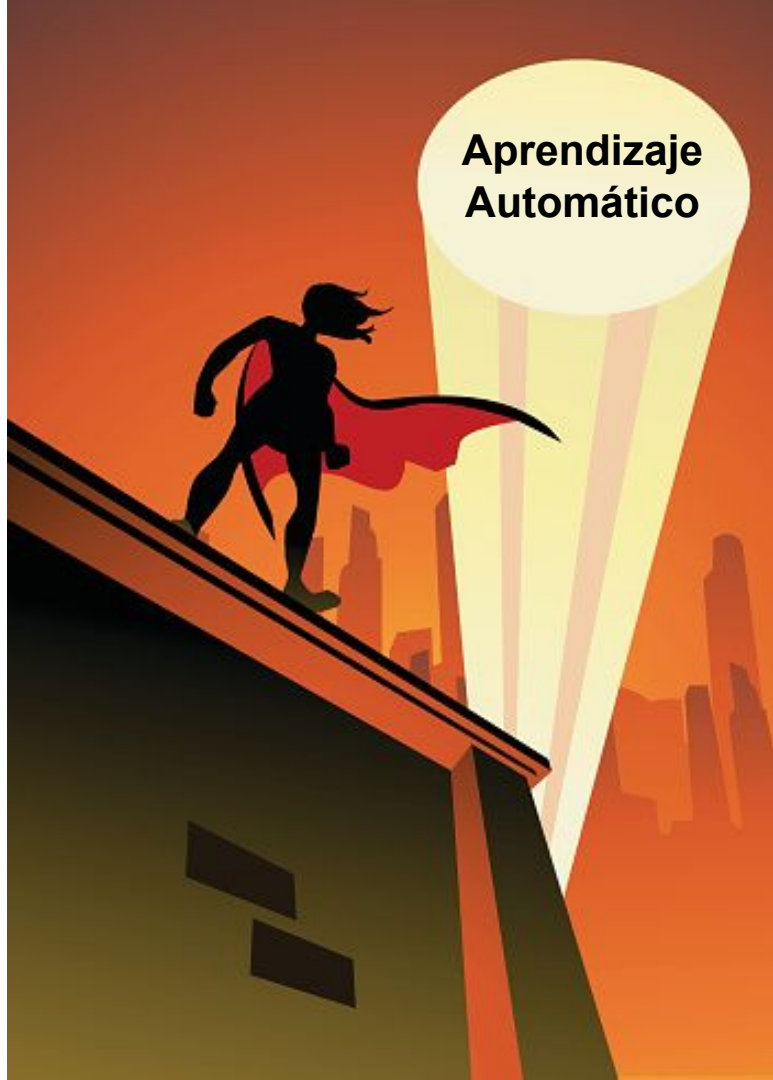
Aprendizaje Automático

Rama de la Inteligencia Artificial, que estudia algoritmos para que las computadoras **aprendan a resolver problemas** a partir del **uso de datos** sin ser programadas específicamente.

- ★ Extraer conocimiento de los datos
- ★ Estadística + AI + Computación + ...
- ★ Cómo construimos programas que automáticamente mejoren en base a la experiencia?

Definición - Mitchell (1998)

Un programa de computadora se dice que aprende de una **experiencia E** con respecto a una clase de **tareas T** y una **medida de performance P**, si su performance en las tareas *T*, medidas por *P*, mejoran con la experiencia *E*.



Aprendizaje Automático

Aplicaciones

Minería de datos (registros médicos, bioinformática, clickstream analysis)

Soluciones que no se pueden programar específicamente (autos autónomos, reconocimiento de escritura a mano, visión)

Procesamiento del lenguaje natural:
Google Translate, ChatGPT

Reconocimiento del habla: Siri, Cortana, Google Now, Alexa

Detección de fraude: bancos, PayPal, Mercado Libre

Publicidad online: Google Ads

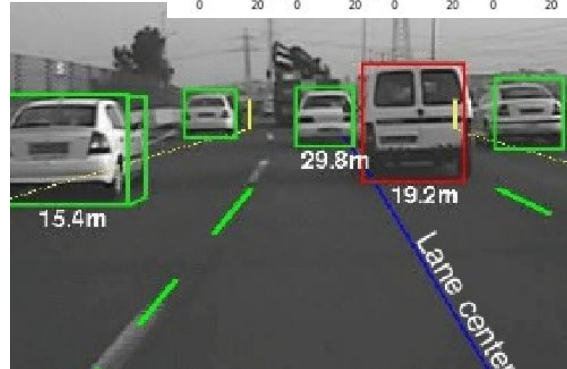
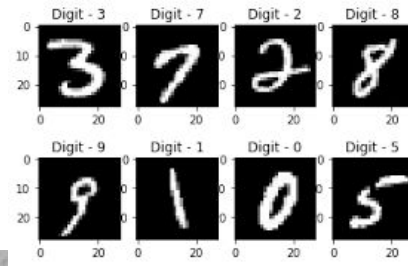
y más ...



Google Ads



ChatGPT



¿Cómo queda esto en términos de la definición del Mitchell? (3min)

Queremos

Definición - Mitchell (1998)

Un programa de computadora se dice que aprende de una **experiencia E** con respecto a una clase de **tareas T** y una **medida de performance P**, si su performance en las tareas *T*, medidas por *P*, mejoran con la experiencia *E*.

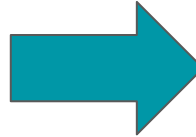
E = Muchas descripciones de perros y gatos

T = Clasificar perros de gatos

P = #Aciertos/#Predicciones

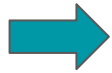
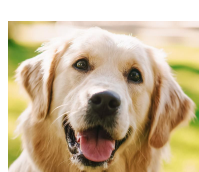


Altura	Ancho de orejas	Largo del hocico	Largo del pelo
100	5	3	Corto

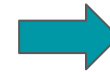
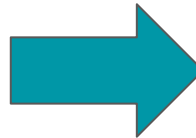


Modelo

Predicción

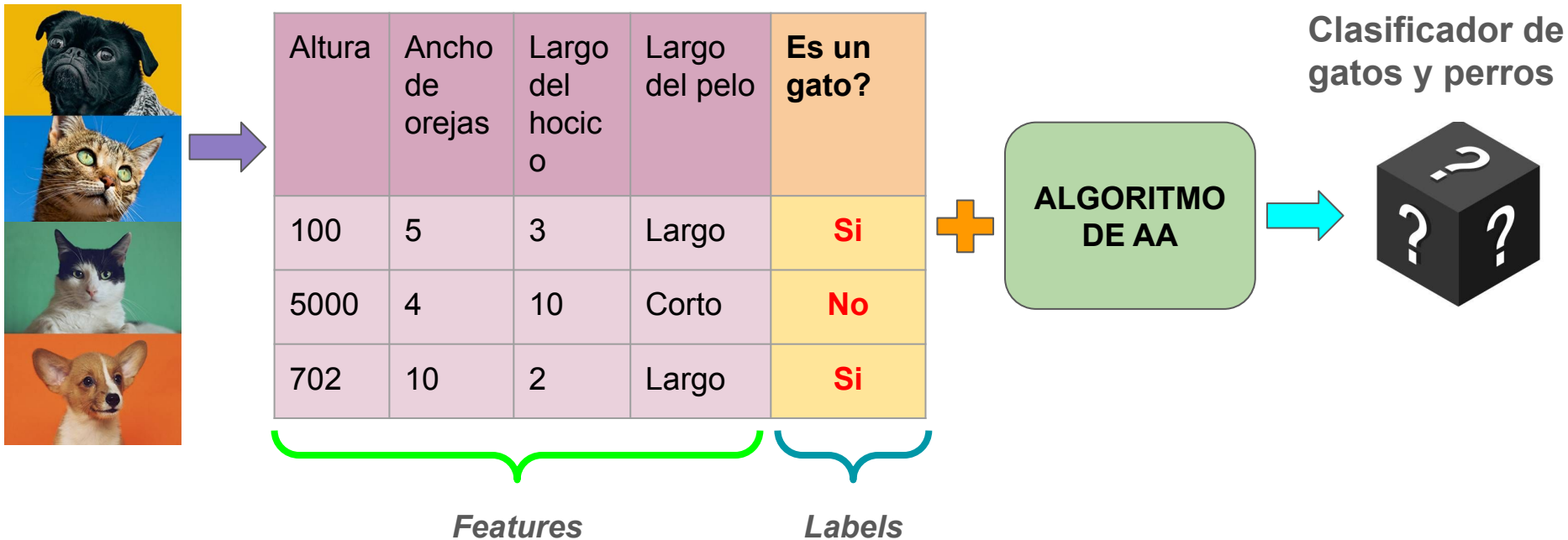


Altura	Ancho de orejas	Largo del hocico	Largo del pelo
5000	4	10	Largo



¿Pero cómo aprende el modelo a hacer lo que queremos?

Podríamos por ejemplo,



Instancia - Una fila del dataframe.
Acá la descripción de un gato/perro

Esto es lo que veremos como
aprendizaje supervisado en 2min

Un poco de metodología...

Qué necesitamos para aplicar AA?

Piensen 5min y lo charlamos

1 Definir la tarea a aprender

- *Cuál es la pregunta a responder? Cómo mido?*

2 Juntar datos

3 Entrenamiento del modelo

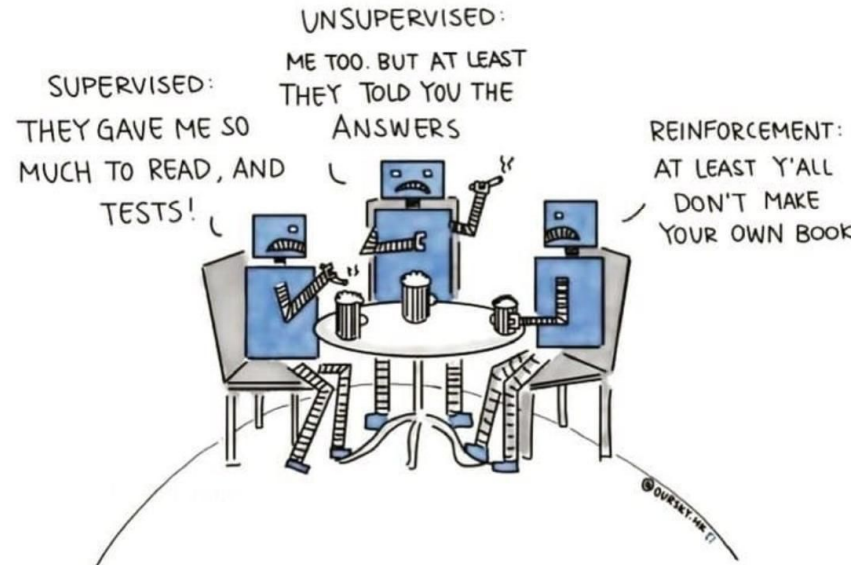
- *Qué modelo uso? Con qué parámetros?*

4 Hacer predicciones!

Cómo anda este modelo “in the wild”?

Esta metodología es ideal para
Aprendizaje Supervisado.

Tipos de Aprendizaje Automático

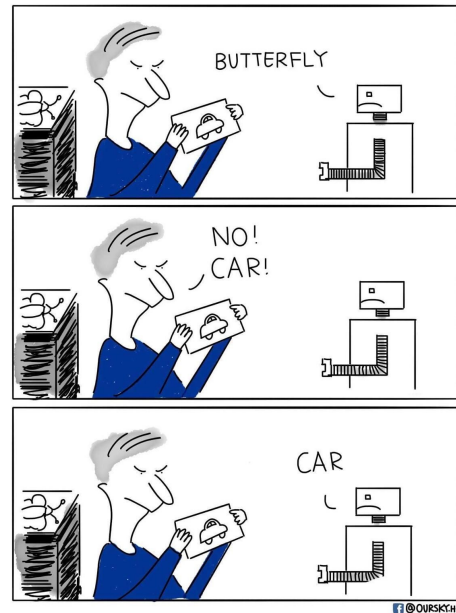
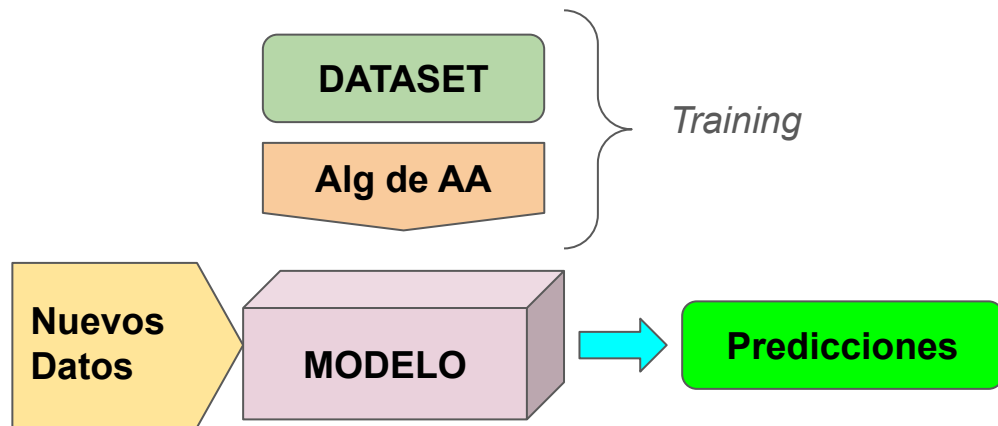


Aprendizaje supervisado

Dados una serie de pares $\{\text{input}(i), \text{output}(i)\}^n$

Se construye un **modelo** que permita crear un output a partir de un input que **nunca vio antes** sin la ayuda de decisiones hardcoded por humanos.

Un algoritmo "aprende" un mapeo que relaciona *input* \rightarrow *output*.



```
if case "miau":  
    return "GATO"  
else:  
    return "NO GATO"
```

El algoritmo de AA genera las reglas por nosotros

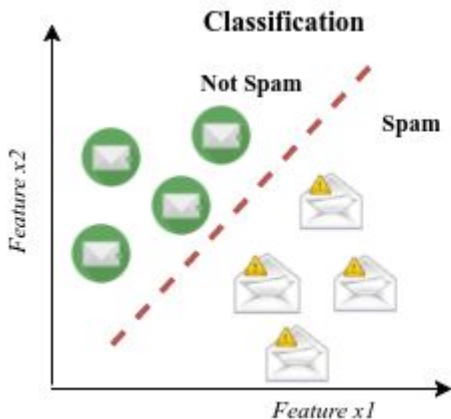
Tipos de problemas

Queremos **aprender a aproximar** una **función desconocida** pero de la cual tenemos ejemplos.

Clasificación $\text{output}(i) :: \text{Bool} \mid \text{Enum}$

Lo que intentamos predecir son un **conjunto de etiquetas, sin orden**.

ej, {GATO, PERRO}, {SOLEADO, LLUVIOSO, NUBLADO}

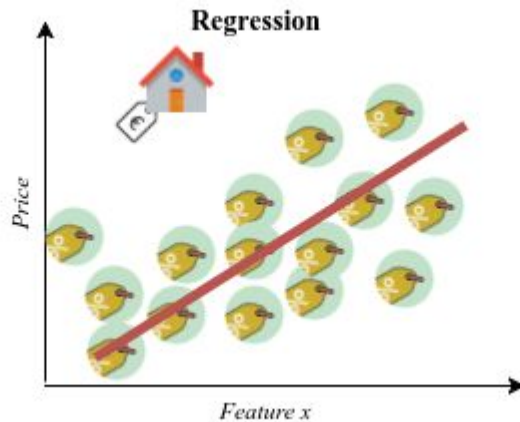


¿Cuál es la pregunta que queremos responder?

¿Es lo mismo predecir los precios de casas que clasificar gatos vs perros? (Piensen la diferencia 3 min)

Regresión $\text{output}(i) :: \mathbb{Z} \mid \mathbb{N} \mid \mathbb{R}$

Lo que intentamos predecir son **valores cuantitativos** ej, Precio de una casa, Sueldo de una persona, Temperatura para mañana, etc.



Piensen un ejemplo de cada uno! (5min)

¿¿¿¿Y los datos????

→ Necesitamos **datos anotados**, es decir con labels.

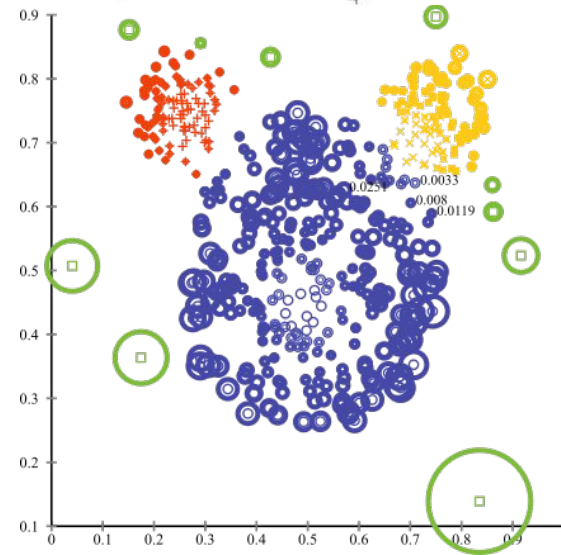
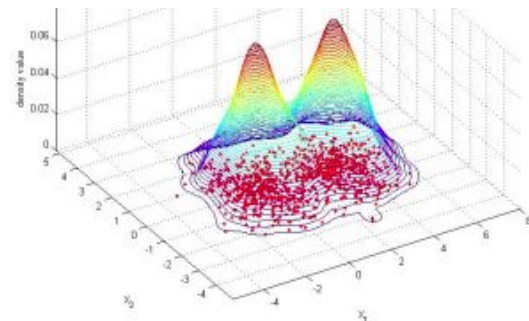
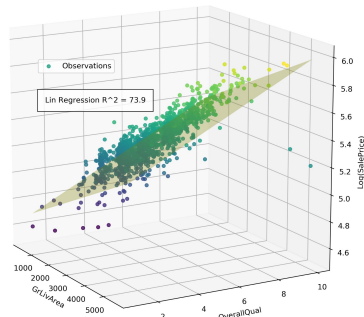
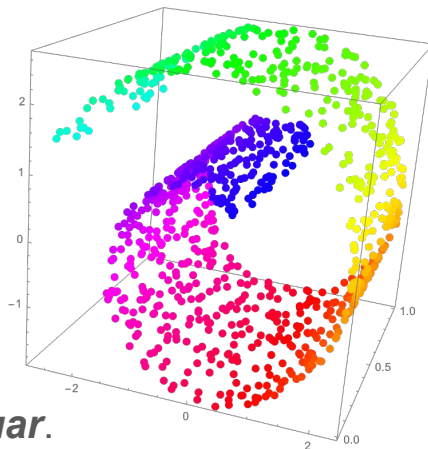
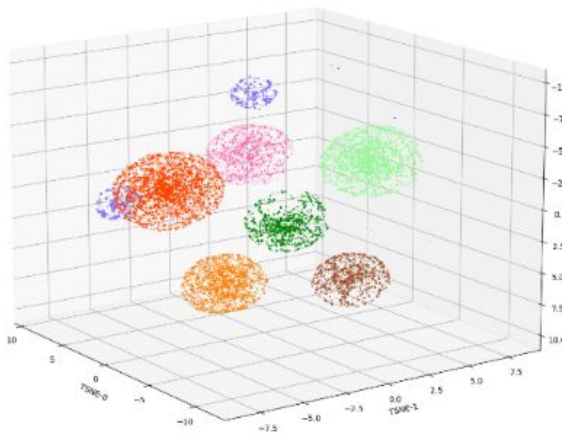
¿Por qué esto es un problema? (discutan 3min)

- ¿Tengo esos datos?? Son caros de conseguir. Se anotan a mano.
- ¿Tengo la suficiente cantidad de datos?
- ¿La representación de los datos me sirve?
- ¿Son datos sensibles? ¿Cuál es la política de protección de estos datos? ¿Los puedo publicar?
- ¿Mis datos me sirven para responder la pregunta que quiero?
- ¿Con cuántos datos entreno el modelo?
- ¿Los datos son una muestra representativa de la realidad?
- ¿Qué métrica puedo usar para medir resultados?



Aprendizaje no supervisado → *No hay labels*

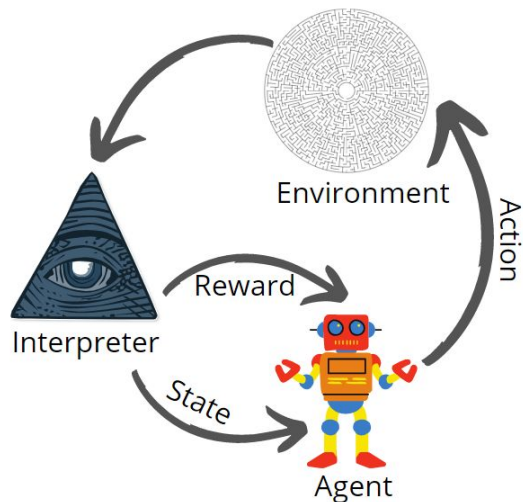
- ★ *Clustering*
- ★ *Detección de anomalías*
- ★ *Reducción de la dimensión*
- ★ *Estimación de densidades*



Estás técnicas suelen ser **difíciles de evaluar**.
Suelen servir para hacer exploración de datos.

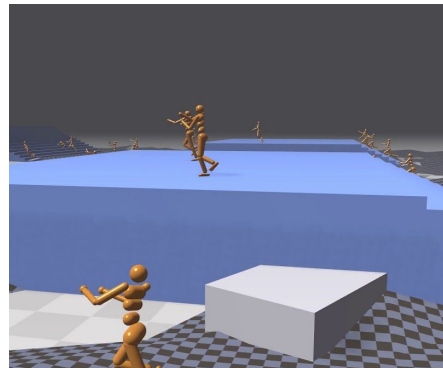
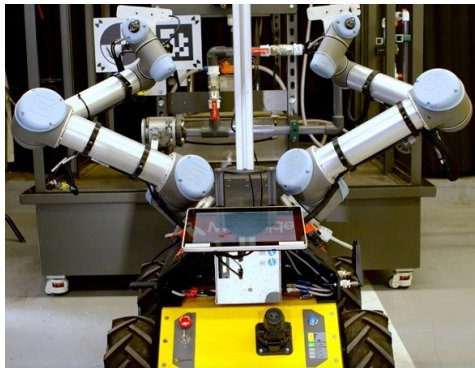
Aprendizaje por refuerzo

- Aprendizaje de un **agente autónomo** (con sensores) para elegir acciones óptimas que le permitan lograr sus objetivos.
- Se usan un sistema de **premio y castigos**, en función de satisfacción del objetivo.
- El agente tiene que aprender qué secuencias de acciones realizar para producir el **mejor premio posible**.



Ejemplos:

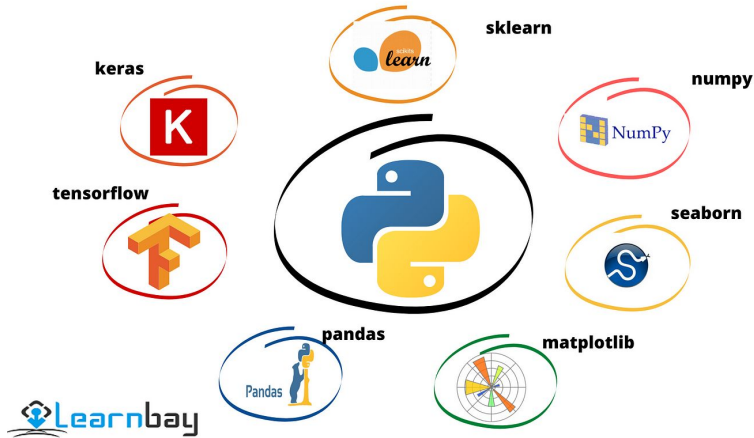
- ★ Agentes que juegan juegos de mesa
- ★ Robots Autónomos
- ★ Simulaciones
- ★ ChatGPT



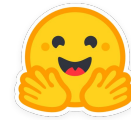
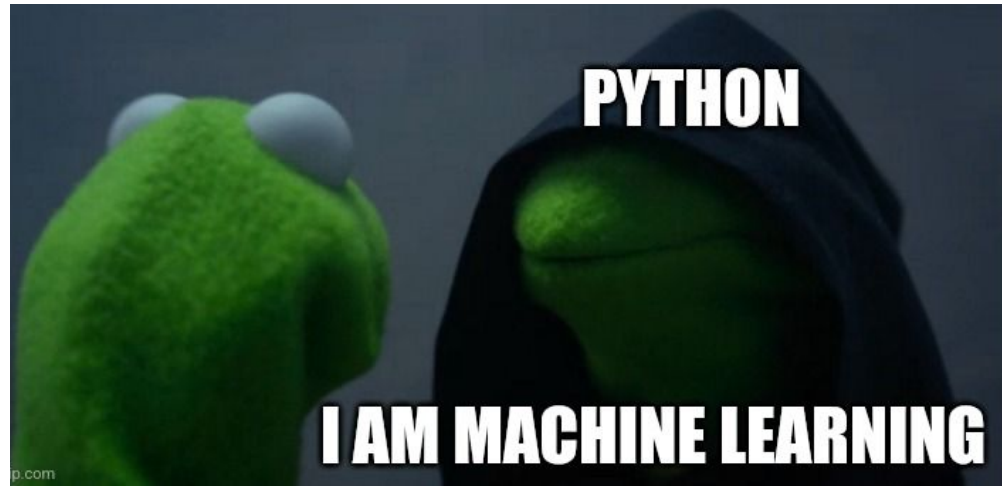
Herramientas



<https://scikit-learn.org/stable/>



<https://scientific-python.org/>



Hugging Face

<https://huggingface.co/spaces>

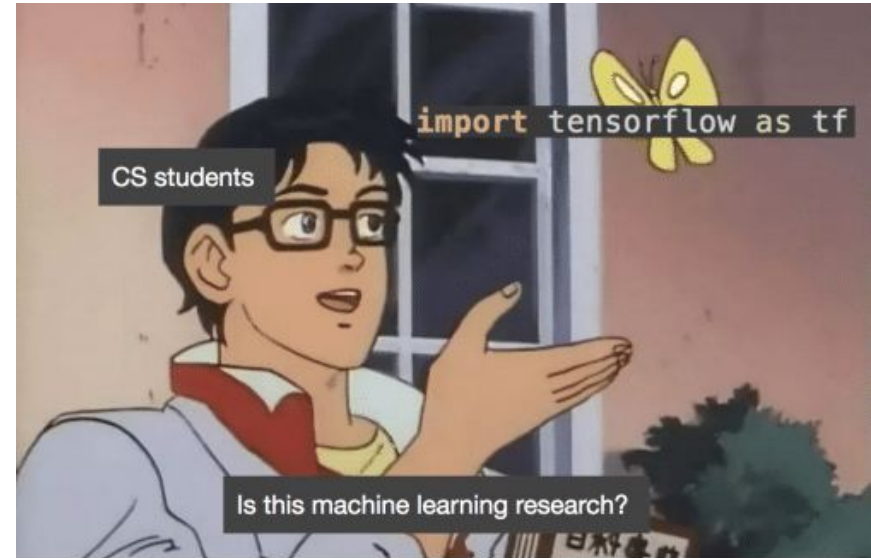
¡Ojo! 🦋

Está buenísimo tener todo esto armado pero está bueno saber **como funcionan** ¡no solo saber usarlas!

La mayoría son proyectos **Open-Source** por lo que ¡está todo el código disponible! (para ver y entender pero también para contribuir)

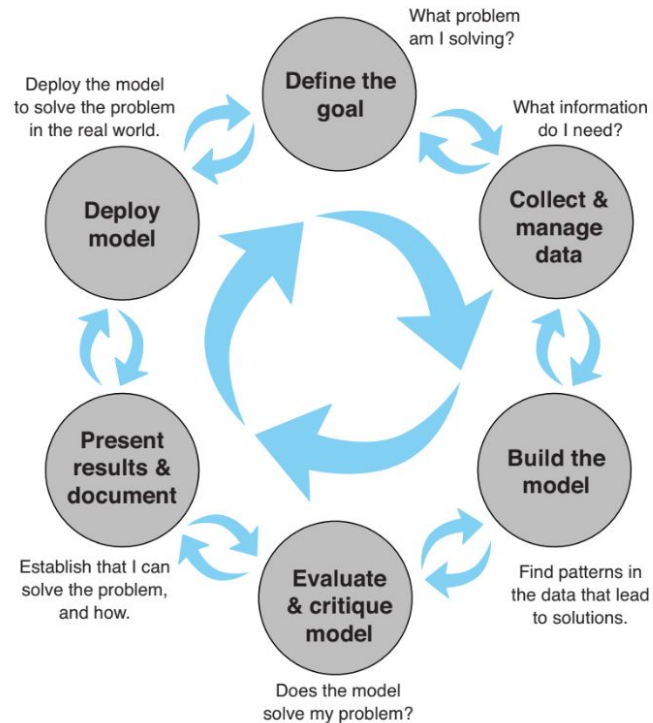
Tarea para la vida:

¡Muchos de los algoritmos que vimos en clase se entienden mucho mejor si nos sentamos a programarlos!



Programmers Nowadays

Posibles etapas de un proyecto de Ciencias de Datos



Resumen

- Aprendizaje Automático vs algoritmos clásicos
- Aplicaciones
- Tipos de Aprendizaje Automático
 - ★ Supervisado (*Clasificación vs. Regresión*)
 - ★ No supervisado
 - ★ Por refuerzo



The background is a vibrant, abstract composition of colors including yellow, green, blue, and purple, with a complex, textured pattern. A central white rectangular box contains the text. Two small black circular marks are visible on the left and right sides of the image, above the text box.

¿¿¿¿Preguntas ????

Bibliografía + Créditos

- “Machine Learning” (Mitchell), Capítulo 1.



Créditos

- Clase 1 de la materia de AA dictada por Pablo Brusco en 1c2023
- Clase correspondiente a este tema del cuatri anterior dictada por Vivi Cotik
- Autores de los magníficos memes e imágenes.